

# 关于高校学位论文数据库建设的思考

杨巧云

(厦门大学图书馆 福建 361005)

**摘要** 分析了学位论文的特点及其开发利用意义,介绍了国内外几个常用的网络型学位论文数据库,并进行了比较分析。在借鉴国外数据库经验的基础上完善我国论文数据库建设;建立有效的学位论文收集机制,注重检索功能和数据库质量建设,最后就进一步扩大论文资源开发提出了几点思路。

**关键词** 学位论文 PQDD CDDDB CDMD CALIS 数据库 信息资源

学位论文(Dissertation for Academic Degree)是高等学校或研究机构的学生为取得学位资格,在导师指导下,参阅大量文献,经反复实验及调研所撰写的科学研究成果的书面报告。它把本学科的基础理论知识与最新发展动态系统地结合在一起,既反映学科发展过程,又提示了未来发展趋势,因此学位论文是了解国内外社会科学、自然科学的研究水平和动态的重要信息资源。

## 1 学位论文的特点分析

### 1.1 内容新颖

学位论文的选题与结论不与他人雷同,具有独到的见解,其研究方向明确、稳定,且内容新颖,有较强的科研继承性和持续性,是在专家教授指导下开展研究的。

### 1.2 信息量大、涉及面广

学位论文比一般性的论文篇幅大、引用的资料详实、数据可靠,阐述问题较为系统、详尽。学位论文的研究范围涉及社会科学、人文科学和自然科学等领域,因此它包括了社会各个领域的新经验、新成果等已知的信息和发展中大量未知的信息。

### 1.3 学术价值高

尽管学位论文的内容深度不同,质量参差不齐,但大多数学位论文是学生在导师指导下完成的研究结晶,这些研究成果有着极高的学术应用价值和参考价值。学生在撰写论文过程中需要查阅大量的国内外文献,这些文献凝聚了一定的文献情报价值;有的学术论文经过反复的实验论证,揭示了一些高新技术的关键所在,因而又具有特定的情报功能。

### 1.4 不公开性

学位论文在通过答辩后,原文绝大多数不公开发表或出版发售,流通范围比较窄,数量极其有限,一般仅限于内部传阅。其原始文献不易获得,有的学位论文具有保密性,所以具有灰色文献的典型特征:时间性和隐蔽性。

## 2 开发学位论文资源的意义

高校图书馆拥有丰富的学位论文资源,对这一文献资源进行开发和利用,具有深刻的意义:

### 2.1 保证教学和科研的连续性

许多课题的研究是导师和连续几届的学生多年钻研的成果,是教与学的结晶,是一个循序渐进的过程。通过图书馆有效开发学位论文文献,有利于保存和提供历史素材,保证教学和科研的连续性,并有助于学生借鉴经验、开阔视野。

### 2.2 有利于学位论文的共享和利用

由于学位论文是一种灰色文献,时效性差、不易流通、原始文献难以获取和利用。通过积极开发学位论文,使原始的论文信息获得升值,有利于文献的传播,达到资源共享的目的,同时也可避免重复选题、避免了科研工作的重复劳动,防止资源浪费。

### 2.3 展示高校图书馆馆藏和服务特色

通过学位论文数据库的建设,能更好地为教学科研服务,并有助于展示高校的科研成果,促进学术交流和成果转化。

## 3 网络学位论文数据库简介

长期以来,由于种种原因,学位论文得不到应有的利用,许多论文被束之高阁,甚至未被收藏保留,造成严重的资源浪费。随着科技发展日新月异,学术交流的日趋活跃,资源共享日显重要,高校学位论文的资源建设与利用已成为业内人士关注的焦点,充分利用网络大力开发共享这类信息资源。目前,网上的学位论文资源越来越丰富,已形成几个较有代表性的学位论文数据库:

### 3.1 博硕士论文数据库(ProQuest Digital Dissertation, 简称PQDD)

它是由美国UMI (University Microfilms International) 公司开发研制,收录了欧美1 000多所大学1861年以来多160多万篇博士、硕士论文。它涵盖的学科几乎包括所有自然科学和社会科学领域,每篇学位论文都可以检索到文摘索引信息。而且该库绝大多数的论文都可以获取到缩微胶片、印刷版或电

子版全文。UMI PQDD是目前世界上最大和使用最广泛的学位论文数据库,它为使用者提供了多种检索方式,而且其信息更新很快,年增长6万条,因此吸引了世界各地用户。

### 3.2 中国学位论文数据库 (Chinese Dissertation Database,简称 CDDB)

中国学位论文数据库资源是由国家法定学位论文收藏机构中国科技信息研究所提供,并委托万方数据加工建库的一个综合性、文献型书目数据库,主要收录了自1977年以来我国自然科学领域博士及硕士研究生论文,其中文摘已近40万篇,首次推出最近3年的论文全文10万多篇,并年增全文3万篇。CDDB是我国最早建设的全国性学位论文数据库,也是我国目前收录学位论文信息最多、最全的数据库,非常适合各所大学及科研机构研究使用。

### 3.3 中国优秀博硕士学位论文全文数据库 (Chinese Doctoral Dissertation and Master's Theses Fulltext Database,简称 CDMD)

它是中国知识基础设施工程(CNKI)的系列产品之一,也是目前国内资源最完备、收录质量最高的博硕士学位论文全文数据库。它收录的论文来源于我国高校、科研院所的博硕士培养点,学科范围覆盖广,以光盘和网络等载体形式向社会提供文献检索咨询服务,经过多年的建设已颇具规模。其结构与《中国期刊全文数据库》基本一致,便于跨库检索、相互链接。

### 3.4 高校学位论文数据库

它是CALIS全国工程文献中心(清华大学图书馆)负责组织,协调进入“211工程”的61所高校合作建设的文献索引数据库。高校学位论文数据库采用统一规范,分散加工,集中建库的运作模式,各个参建单位定期向工程文献中心提交数据。该项目得到全国高校的普遍支持,目前已有近百所高校参加共建,上网数据十几万条,内容涵盖了自然科学、社会科学、医学等各个学科领域,利用率极高,大大推动了高校教学、科研的交流和发展。随后在该库的基础上建设的“高校学位论文全文数据库”,提供了更好的文献服务,全文库采用“通过网上提交方式建立本地学位论文全文库,通过OAI协议集中元数据”的建设模式。

## 4 国内外学位论文数据库比较

随着科技信息工作的不断深入,学位论文数据库已成为高校图书馆自动化建设的重点,各种论文数据库也日趋完善,深受广大用户的青睐。总体上说,我国的学位论文数据库在收录学科范围、记录字段、数据的完整性等方面与国外相比都不逊色,但在

数据量方面,国外数据库远远高于我国,如PQDD数据来源的范围大于我国任何一个数据库,且回溯性好,更新周期也快。

从检索功能设计方面来看,国外数据库有很多地方值得我们学习。以PQDD数据库为例,它提供了基本和高级两种检索方式,检索命中率高,其界面的友好性和易用性都优于我国,体现在:检索字段多、提供的检索手段多。相比之下我国的数据库还难以达到最佳检索效果,并都存在一定的缺陷,如CDDB的专业检索方法有各种符号和规定,比较繁琐复杂;而CALIS库却更适合专业人员使用,缺少再次检索功能。在检索帮助的设置上,国外数据库也比较人性化,可见我国数据库的设计在为用户考虑方面尚不够细致周到。除此之外,国外数据库还充分利用网络的传输功能、互联功能、电子商务功能,来尽量满足用户的需求,PQDD在论文原始文献的提供方面运用了电子商务的方法,每篇论文都有导师链接,而我国的数据库基本上还没有这些功能。由此可见,我国的学位论文数据库建设与国外一些数据库相比较,尚有不足,应加以改进和完善。

## 5 完善论文数据库建设思路

鉴于学位论文具有较高的学术研究价值和实用价值,为了让这些宝贵的知识财富得以充分利用,进一步做好高校论文数据库建设已刻不容缓。虽然目前有的数据库建设卓有成效,但在建设过程中,除认真借鉴国外数据库先进做法外,还应注意以下几方面:

### 5.1 建立有效的学位论文收集机制

由于我国的学位论文不是正式出版物,多年来学位论文的收藏管理一直处于分散状况,所以应引起各高校有关部门的重视,建立起行之有效的呈缴制度,保证学生都能及时上交论文样本;同时还要求学生提交学位论文的电子文本全文,并明确规定电子文档所提交的信息包括个人信息、导师信息和论文信息等,这样才能基本保证论文数据库数据源的质量。除此之外还要加快我国论文数据库的建设效率,努力拓宽论文的数据源。

### 5.2 保护学位论文知识产权

我们开发论文数据库,是为了努力促进文献的传播,达到共享的目的。我们要尊重论文作者的劳动成果,保护其合法的权益。在论文数据库数字化建设中,将作品数字化,这只是一种复制行为,其著作权还应归原著者所享有,所以须征得作者的同意。为了有效保护作者个人的知识产权,同时维护高校的学位论文使用利益,笔者认为各高校应尽早制定“使用学位论文的管理方法”,与作者签订协议,妥善解

决学位论文的知识产权问题。在保护作者著作权的前提下,又要使学校的学位论文得到最大限度的利用,应处理好著作权与资源共享的关系。

### 5.3 完善检索功能,加强用户培训

开展学位论文库的最终目的是应用,在保证数据质量的前提下,可以通过建立合理的检索机制、开发友好的用户检索界面、开展最新文献报导服务等各种手段提高论文库的利用率。因此我们须提供更多的检索点以求完整地揭示论文的内容特征,并设计尽可能多的检索途径以提高查准率、查全率。如提供初级检索与高级检索两种界面,支持二次检索、多种逻辑组合检索等专业检索功能;提供中文简体、中文繁体 and 英文检索三种检索界面。随着全文检索技术、超文本检索技术、模糊检索技术的应用,论文数据库的检索功能也将更加完善、有效。

同时针对用户获取文献信息资源方面存在的问题,还应加强网络信息资源检索和获取能力等方面的培训,保证其能够迅速有效获取各种学位论文数据库的信息资源。

### 5.4 注重论文数据库建设质量

数据的质量是数据库的核心,是一个不可忽视的重要问题。数据的质量主要体现在数据的完整性、数据的准确性和数据的标准化。数据的标准化是网络环境下文献资源共享的前提条件,也是数据发展的基础。为了实现网络条件下的资源共享,所录入的论文数据著录标引必须遵循统一的规范与标准。最后论文数据经过严格的质量检测后才能完成提交,并通过网络提供检索与服务。可见学位论文数据库的联合建库工作是项细致的、专业化程度高的工作,要求建库人员必须掌握扎实的图书情报知识和自动化技术能力,这是决定数据库质量的重要因素,图书馆应创造条件,提供学习机会,提高建库人员的工作能力,保证其能胜任学位论文数据库的建设和开发。

### 5.5 扩大资源开发,提高文献利用率

学位论文是文献信息资源宝库中一个很重要的组成部分,基于Internet这一信息高速通道,我们可以从以下几方面进行论文资源开发,提高文献利用率:

(1)多类型:高校的学位论文库的建设应包括索引库、文摘库、书目库和全文库等多种类型。其中全文库最受欢迎,因此论文数据库最好能提供全文的在线浏览、全文下载、保存、打印等功能,并且也提供摘录功能。

(2)多载体:为适应各种用户需要,学位论文应该提供印刷本、缩微制品、CD-ROM及Web方式联

机等诸多信息服务。

(3)多语种:鼓励论文作者提供多文种版本;图书馆信息开发人员也可精选少数优秀论文翻译成英、法、日等语种,面向世界,扩大影响。

(4)多方式:努力开发学位论文的服务方式,包括印刷本流通、网上包库、镜像站点、局域网光盘、按流量记费等。

(5)多渠道:尽量使论文收入到各种有影响的数据库,扩大传播渠道,从而提高利用率。

此外,还应大力开展馆际协作,通过文献传递系统来满足需要论文全文的用户。目前各种论文数据库的收录年限较短,范围也有限,不能满足各个专业领域用户的要求。所以可尝试采用市场经济的运作机制,促使数据库生产商与高校等论文提交单位签订协议,明确双方的权利和义务,扩大了论文资源的开发,从而更好地完善论文数据库的建设。

### 5.6 妥善处理电子文档

由作者按规范的要求提交论文电子文档,有利于提高论文数据库建设的效率,保证论文及时准确上网,但必须严格做好对电子文档进行的病毒检查及杀毒工作,通过多种杀毒软件检查文档,可以有效地避免病毒感染文档。为了更好地保存电子文档,还可以将电子文档刻成光盘进行保存。

学位论文独特的资源优势已日益得到用户的关注,大力开发学位论文数据库是我国文献机构的一项重要任务。高校图书馆应抓住契机,加强自身的文献资源建设,让学位论文数据库更好地为教学科研服务。

### 参考文献

- 1 刘银红.网络环境下研究生学位论文的开发与利用.图书情报知识,2002(1)
- 2 李淑梅.学位论文资源建设与检索.山东图书馆季刊,2002(2)
- 3 梁孟华,万蜀柏.武汉大学学位论文全文数据库的建设与开发.图书情报知识,2003(3)
- 4 张忠凤.浅议高校图书馆对学位论文资源的开发与利用.河北科技图书苑,2001(6)
- 5 金毅.数字图书馆重要资源——电子化学位论文的发展分析.图书馆建设,2002(4)
- 6 赵美娣,谢敏.中外网络型学位论文数据库比较.图书情报工作,2003(2)
- 7 姚容.中国博士学位论文数据库建设的思考.国家图书馆学刊,2002(3)

(责任编辑:姜雪榕)